

XLSTAT - Analyses factorielles multivariées et classification

3 j (21 heures)

Ref : XSTD

Public

Employé – Technicien – Cadre – Chercheur – Etudiant

Pré-requis

Il est INDISPENSABLE d'avoir des connaissances de base en statistique

Moyens pédagogiques

Formation réalisée en présentiel ou à distance selon la formule retenue
Exposés, cas pratiques, synthèse, assistance post-formation pendant trois mois
Un poste par stagiaire, vidéoprojecteur, support de cours fourni à chaque stagiaire

Modalités de suivi et d'évaluation

Feuille de présence émargée par demi-journée par les stagiaires et le formateur
Exercices de mise en pratique ou quiz de connaissances tout au long de la formation permettant de mesurer la progression des stagiaires
Questionnaire d'évaluation de la satisfaction en fin de stage
Auto-évaluation des acquis de la formation par les stagiaires
Attestation de fin de formation

Objectifs

- Identifier quelle méthode d'analyse multivariée ou de classification utiliser selon le contexte
- Connaître les concepts mathématiques inhérents à ces méthodes
- Mettre en œuvre les différentes analyses
- Interpréter les résultats de chaque méthode
- Analyser les différents graphiques qui en découlent
- Connaître les coefficients permettant d'estimer la qualité de l'analyse statistique
- Mesurer la contribution de chaque variable et de chaque individu sur les axes factoriels

Programme détaillé

PRENDRE EN MAIN L'OUTIL XLSTAT (SI NECESSAIRE)

- Généralités et interface utilisateur
- Interface de base

Rappel sur quelques outils Excel nécessaires à la manipulation XLSTAT.

Activation, chargement et fermeture XLSTAT

Gestion et organisation des données

L'interface XLSTAT

Menus et barre d'outils

Principes de paramétrage des boîtes de dialogue

Gestion des classeurs et des feuilles Excel

Paramétrage de base de l'outil

Présentations des différentes analyses statistiques disponibles

Complémentarités entre Excel et XLSTAT

Outils XLSTAT non statistiques

Repérage de données selon critères

Différents types de fonctionnalités de préparation des données

Regroupement des données en classes

Transformation de données

Outils complémentaires aux graphiques (étiquettes, axes, facteur de zoom,...)

Codage de données

CONNAITRE LES GENERALITES SUR LES METHODES D'ANALYSES FACTORIELLES ET SUR LA CLASSIFICATION

Limites des statistiques classiques

Structure des jeux de données

Champs d'application des méthodes d'analyses multifactorielles

Les objectifs des techniques d'analyse

Objectifs de description

Objectifs de prédiction

Présentation de l'éventail des méthodes

Analyse en composantes principales

Positionnement multidimensionnel

Analyse factorielle des correspondances

Analyse factorielle des correspondances multiples

Analyse canonique des corrélations

Analyse factorielle discriminante

Méthodes de classification : classification ascendante hiérarchique, réallocation dynamique

Les outils mathématiques

Notion de distances

Notion de corrélations

METHODES FACTORIELLES MULTIDIMENSIONNELLES

METTRE EN ŒUVRE L'ACP

Structure du jeu de données et contexte d'application

Matrice de corrélations, de variance-covariance

Inertie et variance

Choix des axes de représentation (choix du nombre de composantes principales)

Interprétation des axes factoriels

Contribution des individus et des variables aux axes

Qualité de représentation des individus et des variables sur les axes

Les différentes ACP : normée et non normée, non paramétrique, polychorique

Représentations graphiques diverses

Positionnement d'une variable illustrative catégorielle et ellipses de confiance

Interprétation des sorties logicielles

METTRE EN ŒUVRE LA MDS

Contexte d'utilisation de la méthode du positionnement multidimensionnel

Structure du jeu de données

Notions de similarité, dissimilarité, distance

Lien et différence entre ACP et MDS

Principes de la MDS

Interprétation des sorties logicielles

METTRE EN ŒUVRE L'AFC

Contexte d'utilisation d'une analyse factorielle des correspondances

Structure du jeu de données

Tableau de contingence

Données individus, variables qualitatives

Différence entre ACP et AFC

Proximités et oppositions entre les modalités, liaison entre deux variables qualitatives

Profils lignes et profils colonnes

Choix des axes de représentation

Représentation des modalités de base autour des composantes

Contributions des modalités à l'inertie

Interprétation des sorties logicielles

METTRE EN ŒUVRE L'AFCM

Contexte d'utilisation d'une analyse factorielle des correspondances multiples

Structure du jeu de données

Différence entre AFC et AFCM

Tableau disjonctif complet, tableau de Burt

Valeurs propres et choix des axes de représentation

Information sur les individus

Rapports de corrélation

Interprétation des sorties logicielles

METTRE EN ŒUVRE L'ACC

Contexte d'utilisation d'une analyse canonique des corrélations
Structure du jeu de données
Analogie avec l'ACP, la régression linéaire et les autres méthodes d'analyse factorielle
Objectifs et principes de l'ACC
Notion de proximité entre 2 groupes de variables quantitatives
Vocabulaire spécifique : variables canoniques, coefficients de corrélation canonique
Représentation des variables et des individus dans les sous-espaces de chaque groupe

METTRE EN ŒUVRE L'AFD (METHODES DE CLASSIFICATION, DE CLASSEMENT)

Contexte d'utilisation de l'analyse factorielle discriminante
Structure du jeu de données
Notions de classement et de discrimination
Méthodologie de l'AFD
Comparaison avec l'ACP
Interprétation des sorties logicielles

METTRE EN ŒUVRE LA CAH

Présentation des objectifs de la classification ascendante hiérarchique
Structure du jeu de données
Choix de la distance, dissimilarité, similarité entre les individus : distance euclidienne...
Choix du linkage (saut, critère de « rapprochement » entre les classes) : méthode de Ward...
Lecture d'un dendrogramme
Choix du nombre de classes
Classification sur les individus
Classification sur les variables (distance de corrélation et composantes latentes)
Interprétation des sorties logicielles

METTRE EN ŒUVRE LA CLASSIFICATION PAR REALLOCATION DYNAMIQUE (CENTRES MOBILES)

Contexte d'utilisation de la classification par réallocation dynamique
Principe général des différentes méthodes
Comparaison avec la CAH
Présentation de la méthode k-means
Les différentes variantes de classification par réallocation dynamique
Interprétation des sorties logicielles
Conseils de mise en œuvre
